Archives nationales du Québec

Étude de faisabilité de la conversion technique des documents informatiques

Rapport final

préparé conjointement par:
les Archives nationales du Québec
Cogni-CASE inc.

Le 25 Janvier 1996

TABLES DES MATIÈRES

1. INTRODUCTION **3**

2. BUT DU RAPPORT **7**

3. SYNTHÈSE DES BANCS D'ESSAI DE LA CONVERSION TECHNIQUE DES DOCUMENTS INFORMATIQUES **8**

3.1. CONTEXTES MINISTÉRIELS DES BANCS D'ESSAI 8
3.2. TECHNOLOGIES/TYPES DES DONNÉES À CONVERTIR 9
3.3. CHOIX DES DONNÉES À CONVERTIR 10
3.3.1. Nature des données *10*
3.3.2. Confidentialité des données *10*
3.3.3. Liens entre les données *11*
3.3.4. Quantité de données *11*
3.4. PROCESSUS DE CONVERSION POUR CES BANCS D'ESSAI 12
3.5. SYNTHÈSE DES RÉSULTATS DE LA CONVERSION 13
3.5.1. Modèle de la structure et du format des données après conversion *13*
3.5.2. Validation des données *14*
3.6. SYNTHÈSE DES IMPACTS DES BANCS D'ESSAI DANS LES M/O. 14
3.7. FAISABILITÉ DE L'EXPLOITATION DES DONNÉES CONVERTIES 16
3.7.1. Outil d'exploitation *16*
3.7.2. Types d'exploitation effectuée sur les données *16*

4. MÉTHODE D'ÉVALUATION ET INDICATEURS DES COÛTS DE CONVERSION **18**

5. EFFICACITÉ ET EFFICIENCE DES OUTILS UTILISÉS **19**

6. SOLUTIONS ADMINISTRATIVES POSSIBLES DE GESTION DES ARCHIVES INFORMATIQUES **21**

6.1. L'EXTRACTION DES DONNÉES 21
6.2. LA CONVERSION DES BANQUES DE DONNÉES 24
6.3. EXPLOITATION DES DONNÉES 25

7. CONCLUSION **27**

8. RECOMMANDATIONS **28**

9. ANNEXE A: LISTE DES MEMBRES PARTICIPANTS **29**

10. ANNEXE B : PROCESSUS DE CONVERSION DES DONNÉES **30**

11. ANNEXE C : SCHÉMA LOGIQUE DU FORMAT UNIQUE **31**

1. INTRODUCTION

La Loi sur les archives (A-21.1) définit un document: «tout support d'information, y compris les données qu'il renferme, lisibles par l'homme ou par machine, à l'exception des documents visés à la Loi sur la Bibliothèque nationale du Québec». La loi doit donc être appliquée par les ministères et les organismes gouvernementaux (M/O) tant pour leurs documents conventionnels sur papier que pour les documents informatiques, c'est-à-dire les fichiers de données ou les banques d'information électronique produits par quelque technologie que ce soit. Les Archives nationales doivent, par contre, être en mesure de conseiller les M/O à l'égard de la gestion de leurs documents informatiques et d'assurer la conservation des documents informatiques de valeur patrimoniale.

Un système de gestion des archives informatiques et le partage des responsabilités entre les intervenants, les ANQ et les M/O, sont à définir entièrement. Toutefois avant d'entreprendre l'élaboration du système de gestion des archives informatiques, les ANQ voulaient vérifier, en collaboration avec le Secrétariat du Conseil du trésor et quelques organismes gouvernementaux, la faisabilité de la conversion technique des documents informatiques, c'est-à-dire les banques de données produites par les systèmes informatiques, sur une technologie unique la plus indépendante possible de l'évolution des technologies de l'information. La problèmatique de la conversion des banques de données devait être élucidée et constituait la première étape à franchir. En effet, advenant l'impossibilité d'envisager la conversion de ces documents informatiques, les orientations administratives actuelles des ANQ concernant la gestion des archives seraient à redéfinir à l'égard de ces documents informatiques et par voie de conséquence le partage des responsabilités avec les M/O serait également affecté. De plus le versement, par les organismes gouvernementaux, de documents informatiques de différentes technologies serait exclu car il serait hors de question de constituer un musée de la technologie pour assurer la diffusion de ces archives.

Afin de résoudre cette problématique de la conversion de la banque de données, les ANQ ont procédé au lancement d'un appel d'offres pour la réalisation d'une «Étude de faisabilité de la conversion technique des documents informatiques». La firme Cogni-Case inc. a été retenue.

Les objectifs poursuivis par ce projet étaient de vérifier la faisabilité technique de conserver centralement le patrimoine documentaire informatique des Québécois, c'est-à-dire :

- identifier les moyens de convertir pour conserver les banques de données structurées et produites par les systèmes informatiques;

- démontrer que ce patrimoine peut être accessible aux chercheurs et au grand public en général.

Note: Les autres types de documents produits à l'aide d'ordinateurs seront considérés lors de l'élaboration de la conception administrative.

Prenant en compte la quantité de données informatisées, la diversité des technologies, la non-portabilité des données d'un environnement à un autre et la complexité technologique de ce type de documents informatiques très liés aux logiciels qui les ont conçus, Cogni-CASE devait, selon le mandat original, démontrer la faisabilité des points suivants:

- la conservation technique des documents informatiques existant sur les plates-formes maxi, mini et micro-ordinateurs vers une technologie unique la plus indépendante possible de l'évolution technologique;

- la conversion automatisée, à l'aide d'outils déjà existants, des différents types technologiques de banques de données (réseau, relationnel, pseudo relationnel, hiérarchique) vers un format séquentiel unique;

- la conservation, dans les fichiers de données converties, des liens entre les données de la banque d'information originale;

- le maintien, à la suite de la conversion technique, de la cohérence et de l'intégrité des données;

- la conversion de la définition des données;

- la possibilité de lire et d'exploiter les données converties selon une approche simple pour un chercheur et pouvant être supportée par l'architecture technologique des ANQ;

- la possibilité de convertir, pour chacun des types technologiques mentionnés, certains regroupements de données, donc des conversions/extractions partielles des banques d'information électronique tout en maintenant la cohérence des données extraites et la possibilité de les exploiter pour des fins de recherche; cet essai a pour but de démontrer que si certaines données d'une banque d'information ont une valeur patrimoniale, il pourrait être possible de conserver seulement les données de valeur patrimoniale et non la banque d'information entière. La différence de coût entre l'option conversion de la totalité ou d'une partie de la banque d'information doit être documentée.

Les biens livrables demandés étaient:

- un manuel d'organisation de projet;deux rapports d'étape démontrant la faisabilité des points susmentionnés, dont l'un regroupait les

résultats des trois premiers bancs d'essai et l'autre les quatre suivants;

- un rapport final contenant:

  - une synthèse des expériences de conversion;

  - les impacts dans les M/O des bancs d'essai;

  - les impacts de l'impossibilité de convertir les dictionnaires de données;

  - une méthode d'évaluation et des indicateurs nécessaires à la détermination des coûts de conversion des documents informatiques;

  - la démonstration, sans effectuer de bancs d'essai, que la méthode et les outils utilisés assurent aux ANQ la réalisation, au meilleur coût, des activités prévues dans ce mandat;

  - une ou des solutions administratives possibles de conversion des documents informatiques et les avantages et inconvénients de chacune de celles-ci.

Pour assurer la gestion de ce projet gouvernemental réalisé en collaboration avec le Secrétariat du Conseil du trésor et différents ministères et organismes gouvernementaux, deux comités, dont l'un professionnel et l'autre consultatif, ont été mis en place (voir annexe A). Les bancs d'essai de conversion des banques de données se sont déroulés dans les organismes gouvernementaux suivants:

. Ministère de la Santé et des Services sociaux (MSSS)
. Ministère des Affaires internationnales de l'Immigration et des Communautés culturelles (MAIICC
. Régie de l'Assurance maladie) du Québec (RAMQ)
. Ministère de l'Agriculture des Pêches et de l'Alimentation du Québec (MAPAQ)
. Société de l'Assurance automobile du Québec (SAAQ)
. Ministère de la Culture et des Communications (MCC)
. Ministère de l'Éducation du Québec (MEQ)

Le processus de validation et d'approbation des bancs d'essai effectués dans les organismes gouvernementaux a été assumé par les deux comités susmentionnés:

- le Comité des professionnels a participé à la préparation des environnements d'essai et à la réalisation des bancs d'essai. Il s'est assuré de l'atteinte des objectifs du projet pilote. Les résultats des bancs d'essai ont été validés par la chargée de projet des ANQ, le professionnel du ministère ou de l'organisme concerné et le professionnel du Secrétariat du Conseil du trésor. Il a recommandé tous les biens livrables au Comité consultatif;

- le Comité consultatif a approuvé le démarrage et la planification du projet. Il a désigné les personnnes ressources requises et a déterminé les contraintes à respecter. Il a approuvé les biens livrables et doit recommander le présent rapport aux autorités gouvernementales.

2. BUT DU RAPPORT

Ce rapport a pour but, premièrement, d'obtenir l'approbation, par le Comité consultatif, de l'ensemble des résultats de l'étude de faisabilité de la conversion des banques de données structurées et produites par les systèmes informatiques. Il vise aussi à informer les autorités des ANQ, du ministère de la Culture et des Communications, du Secrétariat du Conseil du trésor et des ministères et des organismes qui ont participé à cette étude.

3. SYNTHÈSE DES BANCS D'ESSAI DE LA CONVERSION TECHNIQUE DES DOCUMENTS INFORMATIQUES

Dans cette section, on retrouve une description sommaire des différents points couverts par les bancs d'essai ainsi que les commentaires s'y rapportant. Deux rapports d'étape ont été produits sur le déroulement des bancs d'essai et sont disponibles, sur demande, auprès de la responsable du projet des ANQ.

3.1. Contextes ministériels des bancs d'essai

Sur les sept(7) bancs d'essai effectués, cinq(5) se sont déroulés dans les environnements de développement et deux(2) dans les environnements de production. Pour ceux-ci, un simple déchargement des données a été effectué, celles-ci étant non-nominatives.

Sept programmations spécifiques ont été réalisées pour extraire les données reposant sur les technologies suivantes: QSAM, VSAM, CLIPPER, IMS, DATACOM, IDMS, DB2. Des utilitaires commerciaux ont été utilisés pour extraire les données des technologies: ORACLE sur mini et micro-ordinateur, INGRES, SUPRA.

La description fonctionnelle des données étant supportée par diverses technologies, il s'ensuit qu'un effort de documentation sera requis pour rencontrer les besoins de difusion et d'exploitation future des données. Dans le cadre des bancs d'essai, il était impossible de procéder au transfert automatisé de ces informations. En conséquence, une partie du dictionnaire des données n'a pu être produite par la firme pour la démonstration de l'exploitation des données converties.

Commentaires

Pour les technologies les plus récentes, c'est-à-dire de type relationnel ou pseudo-relationnel, des utilitaires commerciaux ont généralement été utilisés pour effectuer les extractions de données. Pour les technologies plus anciennes de types réseau, hiérarchique et indexé séquentiel, des programmations spéciales ont été requises.

On remarque que l'extraction des données de valeur patrimoniale sera facilitée et moins coûteuse si les données ont été migrées sur

des technologies récentes, c'est-à-dire relationnelle ou pseudo-relationnelle.

Par ailleurs, l'inaccessibilité de la description fonctionnelle des données n'a pas posé de problème pour les besoins du projet pilote. Toutefois, si la décision était prise de conserver centralement les documents informatiques de valeur patrimoniale, la documentation de la description fonctionnelle des données devrait obligatoirement être effectuée par le personnel des M\O, ce qui pourrait représenter une charge de travail plus ou moins considérable selon la diversité des données à conserver.

Cependant ce besoin de documentation devient de plus en plus présent dans les M\O avec la création des Centres d'information qui permettent aux usagers de manipuler leurs données, ce qui devrait donc faciliter l'obtention de ces informations.

3.2. Technologies/Types des données à convertir

Les technologies/types de données à convertir par ministère/organisme visé dans ce rapport sont mentionnés ci-dessous. Le type de données définit la structure des données dans le fichier ou la banque de données.

Ministères/organismes	**Technologies\ Plates-formes**	**Type des Données**
MSSS	VSAM\Maxi	Indexé
	QSAM\Maxi	Séquentiel
	CLIPPER\Micro	Pseudo-relationnel
MAIICC	ORACLE\Mini	Relationnel
RAMQ	DATACOM\Maxi	Pseudo-relationnel
	IMS\Maxi	Hiérarchique
MAPAQ	INGRES\Mini	Relationnel
SAAQ	IDMS\Maxi	Réseau
	DB2\Maxi	Relationnel
MCC	ORACLE\Micro	Relationnel
MEQ	SUPRA\Maxi	Pseudo-relationnel

Commentaires

Les bancs d'essai se sont déroulés sur les trois niveaux de plates-formes technologiques reconnues au gouvernement. Les différents types de données (hiérarchique, réseau, relationnel, pseudo-relationnel, indexé séquentiel) ont été convertis. De plus, les technologies converties sont celles qui sont majoritairement utilisées dans le milieu gouvernemental. *On peut donc conclure qu'il s'avère faisable d'extraire et de convertir les données de valeur patrimoniale contenues sur les documents informatiques existants.*

3.3. Choix des données à convertir

3.3.1. Nature des données

Les ANQ avaient demandé d'utiliser, de préférence, des données opérationnelles pour effectuer les essais. En effet, ce sont ces données, si elles sont reconnues de valeur patrimoniale, qui ont la plus grande probabilité de faire l'objet de conservation sur support magnétique en raison de leur volume et de leurs nombreuses possibilités d'exploitation par les chercheurs.

Tous les ministères et organismes, sauf un, ont fourni des données opérationnelles; l'autre a fourni des données administratives non-nominatives.

Commentaires

Les données utilisées pour les essais sont représentatives des données qui seraient à conserver pour leur valeur patrimoniale.

3.3.2. Confidentialité des données

Pour respecter la Loi sur l'accès à l'information, des données non-nominatives (de test ou de production) ont été utilisées. Dans un cas, les données nominatives ont été blanchies.

Commentaires

Le respect de la confidentialité des données n'était pas problématique pour effectuer les essais. Toutefois, cet aspect doit faire l'objet d'un examen plus approfondi lors de la conception administrative. En effet, l'informatique permet de masquer les données nominatives et de donner accès au reste du document, ce qui n'était pas possible avec les documents conventionnels. Les règles d'accessibilité aux documents de valeur patrimoniale devraient être revues pour prendre avantage des possibilités technologiques de protection de la confidentialité de certaines données d'un document informatique.

3.3.3. Liens entre les données

Quel que soit le type de banques de données converties, les liens entre les données ont été maintenus. Cette caractéristique était importante dans le cadre des bancs d'essai car elle permettait de s'assurer que l'on conservait les mêmes possibilités d'exploitation pour les données converties.

Commentaires

Les nouvelles technologies relationnelles sont les plus faciles à convertir et en particulier en ce qui concerne le maintien des liens puisque ces derniers se répètent d'une table à l'autre.

Ce type technologique étant de plus en plus utilisé dans les nouveaux développements de systèmes informatiques, ceci facilitera d'autant la récupération des données de valeur patrimoniale. Enfin, les structures réseau (IDMS) et hiérarchique (IMS) ont été les plus complexes à convertir.

3.3.4. Quantité de données

Le volume d'enregistrements de données extraites a été très variable. Généralement, il s'est agi d'extraction partielle d'une banque de données. Au MCC, la banque de données pour la région de Rimouski a été convertie dans son ensemble.

L'extraction partielle d'une banque de données s'est effectuée en sélectionnant quelques-uns des segments ou des tables appartenant à une banque de données et en n'extrayant que les informations contenues dans ces segments ou ces tables tout en conservant l'intégrité de ces données.

Commentaires

Le nombre d'enregistrements à convertir n'a pas suscité de problème dans les limites du banc d'essai. Des tests de volume n'ont pas été effectués et devront l'être, si requis, selon les orientations qui seront retenues. Malgré l'extraction partielle de banques de données, les liens entre les données ont été très bien maintenus.

3.4. Processus de conversion pour ces bancs d'essai

Le processus de conversion pour ces bancs d'essai est illustré à l'annexe B. En général, les extractions ont été effectuées dans les M\O et les conversions chez Cogni-Case. Pour les banques de données de type relationnel, un déchargement des données a été effectué dans les M\O; l'extraction s'est déroulée chez Cogni-Case sauf pour la banque de données DB2.

Lors de la conversion, les principales manipulations de données portaient sur la transformation de champs binaire, décimal condensé ou décimal flottant en champs numériques; on ajoutait, lorsque nécessaire, des champs "clé" afin de conserver les liens déjà existants entre les données.

La conversion complète s'est effectuée dans trois M\O: MAPAQ, SAAQ, MEQ. La définition des données résultant des fichiers convertis a été utilisée pour alimenter l'outil d'exploitation des données converties sous un format unique.

M/O	**Extraction**	**Exécution du prog. de conversion**
MSSS:QSAM/VSAM	Au ministère	Cogni-Case
MSSS:CLIPPER	Au ministère	Au ministère
MAIICC	Au ministère/chez Cogni-Case	Cogni-Case
RAMQ	À la Régie	Cogni-Case
MAPAQ	Au ministère	Au ministère
SAAQ(IDMS)	À la Société	Àla Société
SAAQ(DB2)	À la Société	À la Société
MCC	Au ministère/chez Cogni-Case	Cogni-Case
MEQ	Au ministère	Au ministère

Commentaires

Le processus de conversion s'est déroulé soit dans les M/O ou chez Cogni-CASE. Il a été remarqué que ce n'est pas le choix de l'endroit où se déroule la conversion qui influence le plus la charge requise par les M/O, mais le temps de préparation des essais et le respect des normes de sécurité d'accès aux ressources technologiques.

De plus, le processus de conversion ne s'est pas déroulé d'une façon aussi automatisée que le permettraient éventuellement les outils de la firme, ceux-ci n'étant pas tous opérationnels présentement. Dans le cadre de cette étude, l'outil de génération des descriptions de fichiers et l'outil de génération des clés pour les banques IDMS ont été partiellement utilisés; des programmations spéciales ont été effectuées car tous les outils nécessaires n'étaient pas disponibles.

Cependant il faut reconnaître que pas moins de onze technologies différentes ont été converties sur un échéancier très court et avec un budget très restreint. Il en ressort que l'utilisation d'outils automatisés de conversion réduirait les efforts, surtout lorsque des manipulations importantes devraient être réalisées sur les données.

3.5. Synthèse des résultats de la conversion

3.5.1. Modèle de la structure et du format des données après conversion

Le modèle théorique de la structure et du format de conversion est joint en annexe C.

La définition physique des données a été saisie manuellement, à l'exception des définitions provenant des technologies ORACLE et IDMS. La variété des dictionnaires et le budget restreint n'ont pas permis de mettre au point tous les outils d'extraction/ conversion.

De plus, un dictionnaire complet des données converties, incluant la description fonctionnelle, n'a pas été élaboré; cette dernière n'a pas été rendue disponible par les M/O en accord avec les ANQ à cause de la charge de travail requise.

Commentaires

De l'avis de la firme et des professionnels en gestion des données, la conversion des dictionnaires physiques de données est faisable mais devrait être précédée d'une analyse coûts/bénifices afin de déterminer si cette conversion sera automatisée ou manuelle. Une telle étude serait aussi nécessaire pour le transfert de la description fonctionnelle des données.

3.5.2. Validation des données.

La validation des données converties a été réalisée par les responsables de données dans les M/O. Elle s'est effectuée par échantillonnage en s'assurant que les résultats étaient probants. Le contenu et les liens entre les données ainsi que le nombre total d'enregistrements ont été vérifiés.

Commentaires

Une très bonne connaissance de la structure de la banque de données d'origine et de son contenu était nécessaire pour valider les résultats de la conversion des données.

3.6. Synthèse des impacts des bancs d'essai dans les M/O.

Les ressources impliquées sont humaines, financières (coût d'utilisation de l'ordinateur) et documentaires.

Pour trois M/O, la charge de travail a été de 1 à 2 jours/personnes (banques de données relationnelles). Pour les quatre autres, la charge de travail a varié de 7 à 15 jours/personnes dépendamment des travaux nécessaires à la préparation des essais ou des problèmes techniques survenus lors des essais.

Les coûts à l'ordinateur central ont été identifiés pour le seul ministère qui loue des services de traitement informatique; ils ont été de 250$.

Les ressources documentaires fournies ont été la définition des données sur papier et sur disquette, incluant, dans certains cas, les modèles de données.

Commentaires

La charge de travail impliquée pour réaliser les bancs d'essai n'est pas un indicateur de la charge qui pourrait être requise en mode opérationnel. Différentes raisons expliquent cette conclusion.

Premièrement, les M/O ont participé à cette étude en demandant aux ANQ de rendre cette opération la plus transparente possible à leurs opérations régulières et en exigeant peu d'efforts, d'où la non-réalisation de la description fonctionnelle des données par les M/O.

Deuxièmement, les besoins du projet pilote étaient ponctuels; en conséquence, la documentation et les outils utilisés visaient à répondre à ce type de besoins. Des programmations ad hoc ont donc été réalisées et non documentées.

Troisièmement, les intervenants qui ont été impliqués dans le projet pilote ont été réduits au minimum par la nature même des bancs d'essai qui visait uniquement la démonstration de la faisabilité technique. Aussi, des réprésentants des directions informatiques ont été les seules personnes interpellées. En mode opérationnel, les producteurs de documents informatiques (les directions opérationnelles), les responsables de la gestion des documents et les archivistes auraient eu à intervenir dans le processus.

3.7. Faisabilité de l'exploitation des données converties

3.7.1. Outil d'exploitation

Lors de ce banc d'essai, on a démontré la faisabilité de l'exploitation des données qui ont été converties en fichier séquentiel selon un format universel. Pour réaliser cet essai, les fichiers de données de la SAAQ(IDMS), de la RAMQ(IMS et DATACOM) et du MEQ(SUPRA) ont été utilisés. La faisabilité de l'exploitation des données a porté sur l'accès et la manipulation possible des données.

L'outil d'exploitation des données, développé spécifiquement pour les besoins de ce banc d'essai, ne visait pas à démontrer la performance d'un outil technique de manipulation, ni la convivialité de cet outil tant pour les archivistes que pour les chercheurs, mais les possibilités techniques d'accès et de manipulation des données.

Le choix d'un outil approprié aux besoins d'exploitation des données sera considéré lors de la réalisation de la conception administrative.

3.7.2. Types d'exploitation effectuée sur les données

Dans le cadre de la démonstration, on a validé la faisabilité de l'exploitation des données en réalisant les opérations suivantes:

- affichage de la description de certains fichiers selon le format universel; ceci permettait de visualiser pour chaque champ contenu sur le fichier son nom, son positionnement sur le fichier, son format(numérique ou alphabétique) et un indicateur nous permettant de reconnaître les clés;

- affichage du contenu d'un ou plusieurs enregistrements d'un fichier converti, c'est à dire originant d'une base de données IMS, IDMS, DATACOM ou SUPRA, à partir d'un critère de sélection; pour mieux visualiser, un en-tête identifiant le nom de chaque champ était aussi affiché;

- navigation entre les fichiers convertis à partir d'essais préparés par les professionnels du MEQ et de la SAAQ; affichage du contenu de ces records.

Commentaires

Il était important de démontrer au cours de ce mandat que les données converties pourraient être exploitées (lecture, extraction, etc.) par la clientèle des ANQ. En effet, celles-ci ont la responsabilité de conserver les archives mais aussi de les diffuser. Un pas énorme a été franchi, à savoir que les données pourront être conservées et diffusées parce qu'elles seront restructurées sur un format unique en séquentiel. Ceci facilite énormément l'accès aux données archivées car le minimum de connaissances technologiques seront requises puisqu'elles sont les plus simples (exploitation d'un fichier séquentiel), les plus indépendantes de l'évolution technologique (toute dépendance à l'égard du système de gestion de banques de données ayant été outrepassée) et définies de manière constante (un format unique).

Au cours des phases ultérieures du projet, les besoins d'exploitation des données devront être définis en fonction des principaux intéressés afin de leur fournir un outil d'exploitation le plus convivial possible.

4. MÉTHODE D'ÉVALUATION ET INDICATEURS DES COÛTS DE CONVERSION

Les commentaires émis sur la «Synthèse des impacts des bancs d'essai dans les M/O» expliquent qu'il est difficile de faire ressortir de cette étude des indicateurs sur les coûts de conversion.

Par contre, on peut supposer que l'utilisation d'outils automatisés de conversion pourrait réduire les coûts à cause de la transportabilité de ces outils d'un M/O à l'autre pour la conversion d'un même type de technologie et à cause de leur adaptabilité aux changements de structure des banques de données à convertir. De plus, une réduction des coûts de maintenance des programmes de conversion ainsi automatisés serait probable.

5. EFFICACITÉ ET EFFICIENCE DES OUTILS UTILISÉS

L'objectif principal de ce mandat était de démontrer la faisabilité de la conversion de certains types de banques de données les plus utilisés dans le milieu gouvernemental et la possibilité d'exploiter les données converties. La méthode et les outils utilisés pour atteindre cet objectif ont consisté à:

- identifier sept ministères et organismes afin de réaliser des bancs d'essai;
- effectuer la conversion des données pour 11 technologies différentes;
- valider les données converties avec les représentants des M\O, de la chargée de projet des ANQ et du représentant du Secrétariat du Conseil du Trésor;
- exploiter certaines données converties provenant de la Société d'Assurance Automobile du Québec, de la Régie de l'Assurance Maladie du Québec et du Ministère de l'Éducation du Québec;
- produire 2 rapports d'étape portant sur les résultats des bancs d'essai ainsi que sur la démonstration de l'outil d'exploitation des données.

Par cette approche, on a pu recueillir rapidement les commentaires des différents participants provenant des ministères et organismes impliqués. En simulant les différentes étapes du processus qui pourraient être mises en place pour la conversion des documents informatiques, on a pu vérifier la faisabilité de celles-ci et cerner certaines problématiques afférentes.

De plus, la méthode et les outils utilisés se sont avérés efficaces et efficients car ils ont permis de démontrer la faisabilité des points suivants:

- la conversion technique des documents informatiques existant sur les plates-formes maxi, mini, et micro-ordinateurs vers une technologie unique la plus indépendante possible de l'évolution technologique;
- la conversion automatisée ou semi-automatisée à l'aide d'outils déjà existants (IDMS, ORACLE) et la conversion des autres technologies par des programmations spéciales;
- la conversion des différents types de technologies de banques de données (réseau, relationnel, pseudo-relationnel, hiérarchique) vers un format séquentiel;
- le maintien des liens entre les données de la banque d'information originale;
- le maintien, suite à la conversion technique, de la cohérence et de l'intégrité des données même lors d'extraction/conversion partielle;

- la conversion de la définition physique des données pour certaines technologies (IDMS, ORACLE);
- la possibilité de lire et d'exploiter les données converties.

6. SOLUTIONS ADMINISTRATIVES POSSIBLES DE GESTION DES ARCHIVES INFORMATIQUES

En cours de réalisation de l'étude de faisabilité, plusieurs questionnements et réflexions ont été émis sur la concrétisation de la gestion des archives informatiques. Le but de cette section du rapport est d'énoncer ceux-ci et non pas de formuler une ou des solution(s) à privilégier, ce qui devancerait les études à réaliser lors de la conception administrative.

De plus, les commentaires qui sont émis découlent des préoccupations contenues dans l'étude de faisabilité, à savoir: l'extraction des données, la conversion des données et l'exploitation des données. Ces activités, quoique essentielles, ne reflètent qu'une infime partie de tout le processus de traitement des archives et en particulier des archives informatiques.

Enfin, la pertinence de centraliser ou non les archives informatiques n'est pas requestionnée dans ce document. Un des objectifs visés par l'étude de faisabilité de la conversion des documents informatiques était de démontrer que l'on pourrait poursuivre la conservation du patrimoine québéçois tant pour les documents informatiques que pour les documents conventionnels. Les conclusions de cette étude sont à l'effet que, sur un plan strictement technologique, la centralisation, la conservation, le traitement et la diffusion des banques de données complexes et dépendantes des technologies qui les ont produites sont faisables. Celles-ci représentent une partie importante des futures archives informatiques. Ce serait donc en référence à d'autres critères que des critères technologiques qu'un requestionnement sur la centralisation et la conservation de ces archives informatiques pourrait être effectué.

6.1. L'extraction des données

La principale interrogation qui a été soulevée au cours de cette étude est la suivante: est-il préférable d'extraire une banque complète de données dans le M/O et d'effectuer, par la suite, les extractions archivistiques ou inversement de déterminer préalablement les données de valeur patrimoniale et de n'extraire que ces données.

Deux solutions sont donc possibles et commentées sommairement ci-après.

Extraction complète d'une banque de données

C'est l'opération qui pourrait être la plus simple pour les M/O sur le plan technologique.

Cette façon de faire pourrait accélérer le prise de décision sur les délais de conservation des documents informatiques. En effet, une très bonne connaissance des données et des liens entre celles-ci est requise pour pouvoir choisir non seulement les données de valeur patrimoniale, mais aussi celles qui sont intimement reliées à ces dernières et qui sont requises pour que les données extraites soient utiles ou "parlantes". Cette solution permettrait donc aux archivistes de conserver temporairement toutes les données et de pouvoir reconsidérer leurs décisions sur la sauvegarde de certaines données sans avoir à intervenir de nouveau auprès des M/O pour procéder à d'autres extractions.

Les procédures actuelles d'analyse des délais de conservation des documents pourraient aussi être maintenues, certains retards dans l'analyse archivistique ne causant pas de préjudices aux M/O. Enfin, les tâches d'analyse des calendriers et du traitement archivistique pourraient se dérouler en séquence, l'analyse des données s'effectuant de façon itérative et s'approfondissant d'une étape à l'autre. Les consultations nécessaires auprès d'historiens ou de chercheurs seraient également facilitées.

Par contre, plus la variété de données à conserver sera grande, plus laborieuse sera l'étape de description des données par l'organisme. Or, cette documentation est nécessaire pour le traitement archivistique et la diffusion de ces archives. Évidemment, la probabilité d'une plus grande variété des données s'accroît avec le versement de la banque dans sa totalité. Par ailleurs, l'absence de cette description des données obligera, quelque soit la solution retenue, une participation active des gestionnaires de données avec les archivistes afin de permettre à ces derniers une prise de connaissance et une meilleure compréhension du contenu de la banque de données.

Il existe dans certaines organisations de très grosses banques de données. Dans l'éventualité d'une conservation centralisée, le support de transfert de données et le matériel nécessaire pour les relire devront être pris en considération.

Extraction partielle d'une banque des données

Le principal avantage d'une extraction partielle d'une banque de données est la réduction, dans les organismes, de la charge de travail requise pour documenter fonctionnellement les données ou encore d'adapter cette documentation pour rencontrer les besoins archivistiques.

Les données versées pour leur conservation permanente seraient celles dont la valeur archivistique aurait été identifiée. Le volume de données à conserver serait donc moindre, plus facile à gérer et moins coûteux à maintenir.

Par ailleurs, une extraction partielle des données exige que les besoins archivistiques soient ciblés: extraction de certains champs uniquement, réunion d'informations se retrouvant sur plus d'une table, etc..

Par contre, cette solution requiert une plus grande entente entre les organismes concernés sur les échéances de prise de connaissance de la banque de données, son analyse et sur la prise de décision des données à conserver. Les décisions de conservation peuvent être irrévocables dépendant des délais de révision de celles-ci.

Si cette solution demande moins d'efforts de documentation de la part des organismes, elle exige en contrepartie plus d'efforts de développement pour les M/O ou une firme conseil surtout si des manipulations au niveau de la structuration des données étaient effectuées.

6.2. *La conversion des banques de données*

La réalisation des bancs d'essai a permis de faire ressortir, pour la conversion des documents informatiques, deux enjeux que l'on peut formuler ainsi:

- avec quoi doit-on convertir les documents informatiques? Un outil générique ou des programmes spécifiques à chaque document converti;
- où doit-on convertir les documents informatiques?.

La présente étude n'avait pas pour but de choisir un produit mais de démontrer la faisabilité de la conversion technique des banques de données sur un format unique. Les conclusions de cette étude sont claires et précises: il est faisable de convertir les banques de données. L'utilisation d'un outil générique pour la conversion de deux types de banque de données (IDMS, ORACLE) et l'expertise de la firme en ce domaine laissent présager qu'il est possible d'effectuer des conversions automatisées pour différentes technologies.

Toutefois, l'évolution technologique étant très rapide et la concurrence très féroce, il y aura lieu d'effectuer une étude de produits disponibles au cours de la conception administrative. Ainsi, on pourra mieux identifier les avantages et inconvénients des options proposées.

Quant à la deuxième question, l'expérience des bancs d'essai a démontré que des ressources externes aux ministères et organismes ne peuvent travailler sans l'apport de ressources internes (documentation à produire, validation des résultats, etc.). Il faudrait donc évaluer ultérieurement les options administratives possibles de conversion des documents informatiques ainsi que les coûts/bénéfices associés à chacune.

6.3. Exploitation des données

Dans le cadre de ce mandat, on a démontré la faisabilité de l'exploitation des données ainsi que la manipulation possible des données lorsque celles-ci étaient transformées dans un format séquentiel. A partir de cette démonstration, on peut déjà identifier certains besoins qui devront être rencontrés dans l'élaboration de cet outil. Ainsi, l'outil d'exploitation des données pour les archivistes devrait posséder les caractéristiques suivantes:

- un interface usager très convivial pour des utilisateurs;
- un mécanisme de repérage de l'information basée sur la démarche de recherche effectuée par un archiviste ou un chercheur;
- des fonctionnalités pour la saisie de données permettant l'alimentation d'un dictionnaire;
- un dictionnaire permettant de connaître les fichiers provenant de chaque ministère ou organisme, l'année d'extraction et une brève description de son contenu;
- un dictionnaire permettant de connaître toutes les définitions fonctionnelles de chaque donnée;
- un dictionnaire permettant de faire, pour chaque donnée, le lien entre celle-ci et les éléments physiques qui l'incarnent;
- un dictionnaire permettant de connaître l'emplacement sur un fichier de chaque champ physique;
- des facilités pour visionner le contenu d'un ou plusieurs fichiers;
- des facilités au niveau de l'affichage pour pouvoir exploiter des informations contenues sur plusieurs fichiers;
- des facilités pour produire des données agrégées (ex.: nombre de..., somme de., etc) à partir d'un ou plusieurs fichiers;
- des facilités d'extraction partielle des données d'un ou plusieurs fichiers;
- des possibilités pour produire tous résultats sur papier, fichier ou à l'écran avec l'identification des sources d'information d'origine (ex.: nom du fichier, nom du ministère, date d'extraction, description des données fonctionnelles, etc.).

Ces caractéristiques devraient couvrir la majorité des besoins d'exploitation des documents informatisés. Il faudra donc procéder à une analyse de produit pour assumer adéquatement la fonction d'exploitation des données converties ainsi que les modalités de

diffusion de ces archives, localement et/ou via l'autoroute de l'information.

7. CONCLUSION

Cette étude visait à démontrer que certains documents informatiques, c'est-à-dire les banques de données produites par les systèmes informatiques existants, pourraient être convertis selon un format unique, en séquentiel. L'objectif poursuivi était de vérifier l'indépendance possible de ces archives informatiques à l'égard de l'évolution technologique, la faisabilité de conserver les banques de données de façon permanente sans constituer un musée des technologies utilisées pour les produire et la possibilité d'exploiter les banques de données converties. Sept organismes gouvernementaux ont participé à la conversion de banques de données de types réseau, hiérarchique, pseudo-relationnel et relationnel générées sur les trois niveaux de plates-formes technologiques: maxi, mini et micro-ordinateurs.

Tous les bancs d'essai effectués se sont avérés concluants sur la faisabilité de convertir les banques de données. Par contre, la diversité des dictionnaires de données générés par les organismes gouvernementaux et le budget restreint de cette étude n'ont pas permis de statuer sur la faisabilité de la conversion des dictionnaires. Néanmoins, les possibilités d'exploitation des banques de données converties ont été démontrées.

Les banques de données visées par cette étude pilote représentent les documents informatiques les plus complexes, les plus dépendants de la technologie et les plus volumineux générés par les organismes gouvernementaux. Aussi, leur conversion représentait un véritable défi. C'est la raison pour laquelle cette étude a été restreinte à ce type de documents informatiques. La solution de la problématique technologique reliée aux banques de données ne résout pas la problématique générale de l'ensemble des documents informatiques produits au sein du gouvernement du Québec. Elle apporte toutefois la certitude qu'une partie importante du patrimoine documentaire informatique québécois peut être sauvegardée.

8. RECOMMANDATIONS

Afin d'assurer une gestion adéquate de l'ensemble des documents informatiques au gouvernement du Québec et de sauvegarder le patrimoine documentaire informatique, il est recommandé:

- d'établir la problématique générale de la gestion de l'information électronique gouvernementale;
- de poursuivre la conception administrative en élaborant des options de sauvegarde des documents informatiques, en évaluant les coûts/bénéfices et en recommandant la meilleure solution aux autorités gouvernementales.

La participation et l'adhésion des organismes gouvernementaux à la réalisation de la conception administrative et à la solution administrative qui sera retenue seront garantes de l'atteinte des objectifs visées. Aussi, il est recommandé que la communauté gouvernementale soit sensibilisée et informée de la problématique actuelle de sauvegarde des documents informatiques afin d'obtenir leur collaboration. C'est seulement dans ce contexte que les Archives nationales du Québec pourront exercer leur rôle gouvernemental de sauvegarde et de soutien à la gestion des documents informatiques.

9. ANNEXE A: LISTE DES MEMBRES PARTICIPANTS

COMITÉ CONSULTATIF

Représentants des organismes centraux:

Robert Garon, ANQ président
Micheline Bélanger, ANQ chargée de projet
Annette Plante, SCT

Représentants des M/O:

Gaétan Allard, MCC
Andrée Blanchet, MSSS
Andrée Boulianne, SAAQ
Patricia McGrath, MAIICC
Guy Morneau, MEQ
André Roy, MAPAQ
Claude Vaillancourt, RAMQ

COMITÉ DES PROFESSIONNELS

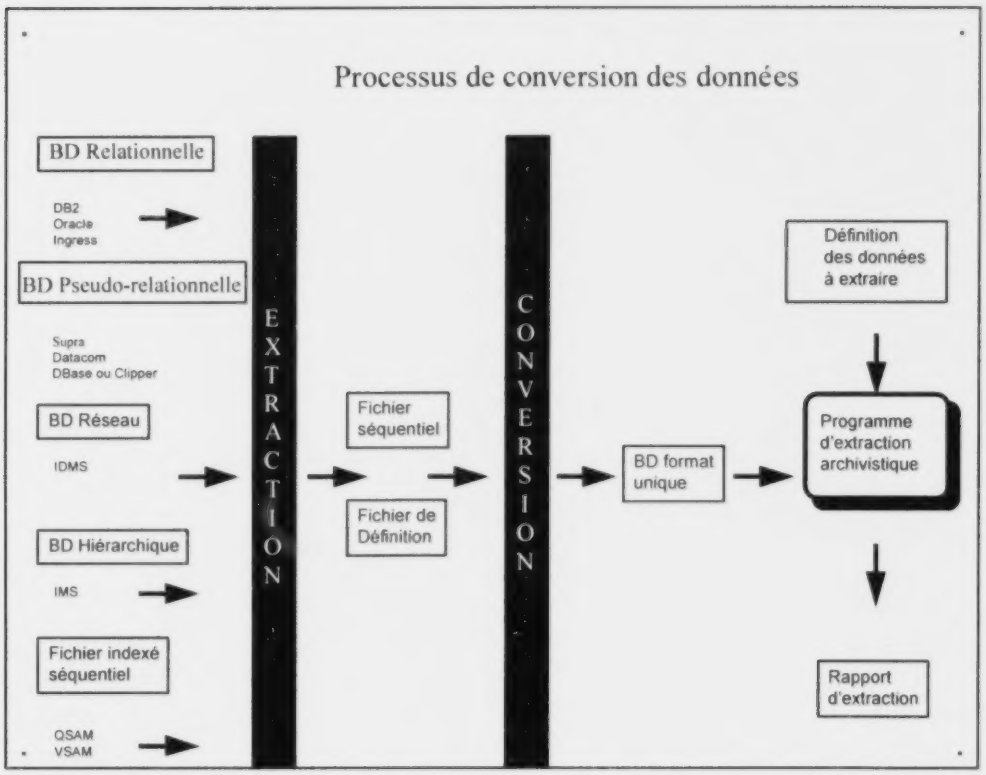
Représentants des organismes centraux:

André Béland, SCT
Micheline Bélanger, ANQ

Représentants des M/O:

Marthe Baril, MEQ
Benoît Boivin, SAAQ
Claude Bouchard, MSSS
Louise Claisse, RAMQ
Christian Gilbert, MAIICC
Daniel Hardy, MAPAQ
Denis Shaink, MCC

10. ANNEXE B : Processus de conversion des données



11. ANNEXE C : Schéma logique du format unique

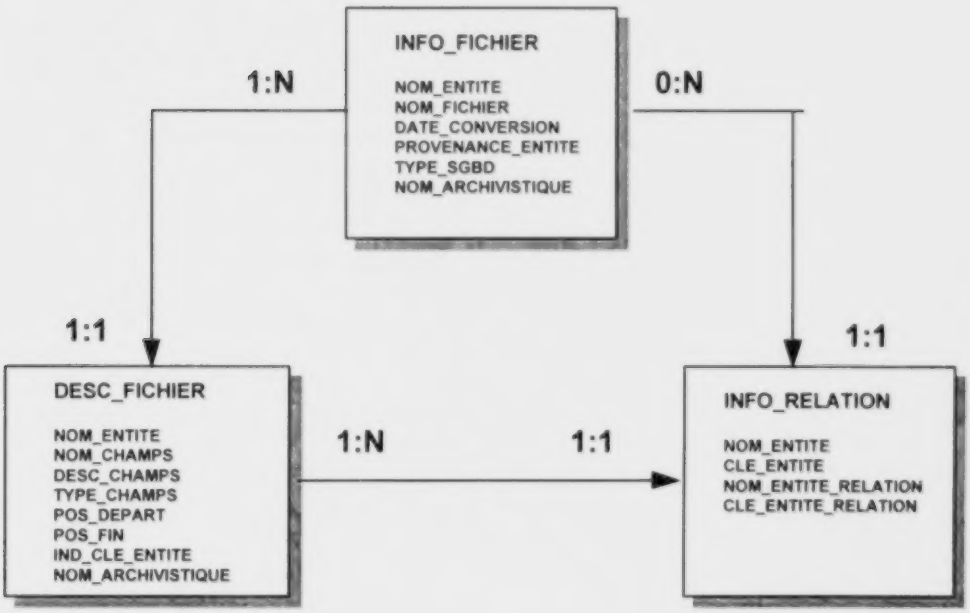



SCHÉMA LOGIQUE DU FORMAT UNIQUE